

How to model mutually exclusive events based on independent causal pathways in Bayesian network models

Norman Fenton¹, Martin Neil², David Lagnado³, William Marsh⁴, Barbaros Yet⁵, and Anthony Constantinou⁶

THIS IS A PRE-PUBLICATION DRAFT OF THE FOLLOWING CITATION:

Fenton, N., Neil, M., Lagnado, D., Marsh, W., Yet, B., and Constantinou, A. (2016). How to model mutually exclusive events based on independent causal pathways in Bayesian network models. *Knowledge-Based Systems, In Press*.

DOI: <http://dx.doi.org/10.1016/j.knosys.2016.09.012>

© 2016. This manuscript version is made available under the CC-BY-NC-ND 4.0 license: <http://creativecommons.org/licenses/by-nc-nd/4.0/>



¹ School of Electronic Engineering and Computer Science, Queen Mary University of London, E1 4NS, UK and Akena Ltd 11 Main Street, Caldecote, Cambridge, CB23 7NU. Tel: +44 20 7882 7860 E-mail: n.fenton@qmul.ac.uk.

² School of Electronic Engineering and Computer Science, Queen Mary University of London, E1 4NS, UK and Akena Ltd E-mail: m.neil@qmul.ac.uk.

³ Dept of Experimental Psychology, University College London, London WC1E 6BT, UK. E-mail: d.lagnado@ucl.ac.uk.

⁴ School of Electronic Engineering and Computer Science, Queen Mary University of London, E1 4NS, UK. E-mail: d.w.r.marsh@qmul.ac.uk.

⁵ Department of Industrial Engineering, Hacettepe University, Turkey. E-mail: barbaros.yet@hacettepe.edu.tr

⁶ School of Electronic Engineering and Computer Science, Queen Mary University of London, E1 4NS, UK. E-mail: a.constantinou@qmul.ac.uk.

Abstract— We show that existing Bayesian network (BN) modelling techniques cannot capture the correct intuitive reasoning in the important case when a set of *mutually exclusive* events need to be modelled as separate nodes instead of states of a single node. A previously proposed ‘solution’, which introduces a simple constraint node that enforces mutual exclusivity, fails to preserve the prior probabilities of the events, while other proposed solutions involve major changes to the original model. We provide a novel and simple solution to this problem that works in all cases where the mutually exclusive nodes have no common ancestors. Our solution uses a special type of constraint and auxiliary node together with formulas for assigning their necessary conditional probability table values. The solution enforces mutual exclusivity between events and preserves their prior probabilities while leaving all original BN nodes unchanged.

Index Terms— Bayesian networks, mutually exclusive events, causes, uncertain reasoning.

1 INTRODUCTION

A Bayesian network (BN) is a graphical probabilistic model that is especially well-suited in decision-making scenarios that require us to consider multiple pieces of uncertain evidence involving causal relationships (Fenton and Neil, 2012). A BN consists of a set of nodes (that represent uncertain variables) and directed edges between those nodes for which there is a causal or evidential relationship. Every node has an associated *conditional probability table* (CPT); for any node without parents the CPT specifies the prior probabilities of each of the node states, while for any node with parents the CPT captures the prior probability of each node state conditioned on each combination of states of the parent nodes. In addition to its powerful visual appeal, a BN has an underlying calculus based on Bayes Theorem that determines the revised probability beliefs of all uncertain variables when any piece of new evidence is presented. This process is called *evidence propagation* (Fenton and Neil, 2012, Pearl, 1988). There are widely available BN tools that implement standard propagation algorithms (see (Murphy, 2014) for extensive list and comparisons), and hence enable non-specialist users to easily build and run BN models. With propagation a BN can be used for both prognostic and diagnostic types of reasoning. In prognostic reasoning we enter evidence about causes in order to reason about effects (we also refer to this as ‘forward inference’) whereas in diagnostic reasoning we enter evidence about effects to reason about causes (we also refer to this as ‘backward inference’).

What we are interested in here is the special case where different possible events or outcomes are necessarily *mutually exclusive* (meaning that only one can be true at any time) but where these outcomes need to be modelled as separate BN nodes rather than states of a single node. We assume that these separate nodes have no common ancestors. In Section 2 we describe why this is a common and important problem and what properties need to be satisfied in any BN that attempts to model mutually exclusive outcomes as separate nodes. Although previous work has touched on the problem (Diez and Druzdzel (2006), Flores et al., 2005, Lam and Yeap, 1992, Jensen and Nielsen, 2007, Pearl, 2000, Pearl, 1988, Perry and Van Allen, 2005) it has never been stated explicitly nor has it been adequately resolved, although the problem of transforming the states of a variable into multiple mutually exclusive variables appears to bear a close resemblance to the problem of transforming an n -ary constraint into multiple binary ones in the field of constraint satisfaction (Samaras and Stergiou, 2005). In Section 3 we review previously proposed solutions and show their limitations. In Section 4 we provide a novel solution to the problem that involves introducing an auxiliary node with

a constraint, and provide the formulas needed to assign values to the new CPTs. Section 5 provides examples and guidelines on where it is appropriate to use the proposed solution.

Executable versions of all of the BN models described in the paper are freely available for inspection and use in the supplementary content.

2 THE PROBLEM

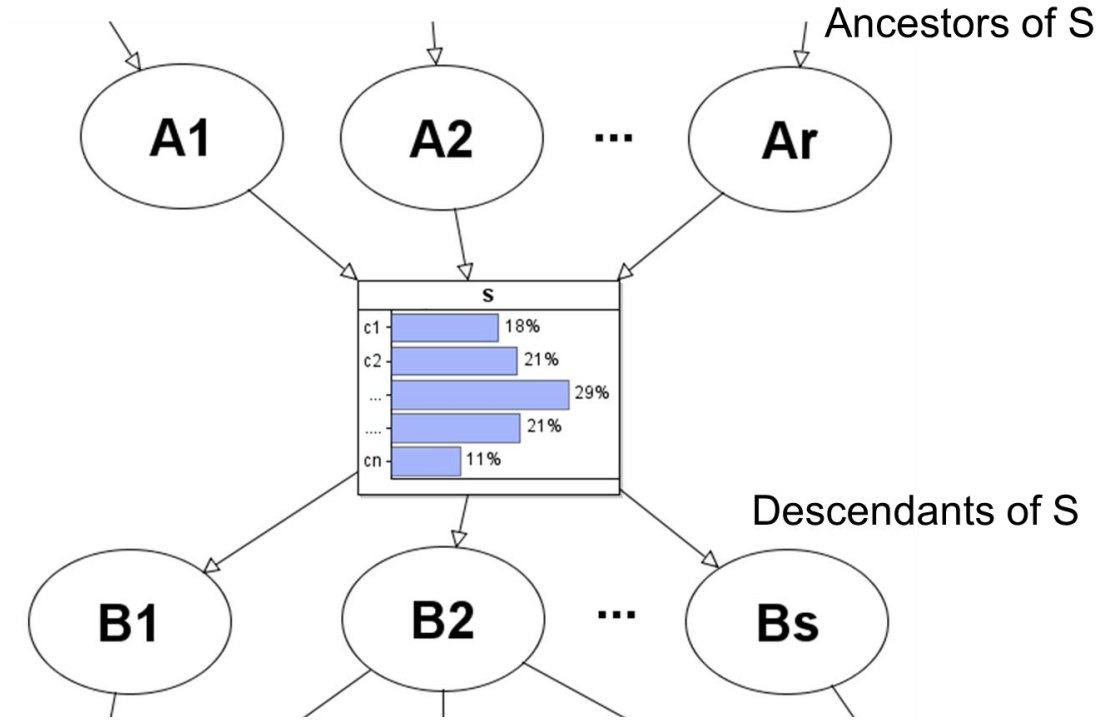


Figure 1 BN model fragment based around a node S with n states

The generic BN in Figure 1 involves a node S – with n discrete states – and a set of ancestor and descendant nodes. This BN structure is typical of many that arise in real world problems, such as in legal arguments and inquests. The states of node S represent n mutually exclusive and exhaustive, but unobservable, hypotheses of which we seek to determine which is/was the most likely. For example, in an autopsy the states of S might correspond to the set of possible causes of death {natural, suicide, accident, murder}. The example in Appendix 1 is of a legal trial of a defendant D where the hypotheses are not simply just {guilty, not guilty}. There are a wide range of applications where the problem occurs and needs to be solved, including any problem in which the events that we wish to determine or predict represent a classification of some outcome. For example:

- Identifying an airborne enemy threat {*Missile, Aircraft, Drone, Other*}:
- Predicting the winner of an election {*candidate1, candidate2, ...candidateN*}

What characterises these sorts of BN model fragments are the following common properties:

- The ancestors of S typically represent separate ‘causal pathways’ for the different states of S . So, in the autopsy case the ‘accident’ hypothesis might involve a narrative with factors such as “participating in dangerous sports”,

while the ‘murder’ hypothesis might involve a narrative with factors such as “in dispute with known criminal gang”

- The descendants of S typically represent diagnostic and other evidence about the individual states of S . For example, evidence of a bullet found in the back of the body supports the murder hypothesis.

The focus of the model is in determining, after observing the evidence and considering all prior assumptions and probabilities, which of the mutually exclusive states of the node S is the most likely. However, there is a fundamental problem in building such a BN: *it requires us to complete CPTs which (in realistic examples with multiple causes and outcomes) are infeasibly large and for which the vast majority of entries are either redundant or meaningless*. For example:

- *The CPT for S* : In the example in Appendix A, even if we assume all causal parents have just two states (true and false), this CPT has $6 \times 2^7 = 768$ entries. Although each causal node influences only one possible hypothesis we are forced to give (redundant) separate probabilities conditioned on every combination of all the other causal factor states. For example, “X often swam in sea” can only influence whether or not “X died accidentally”; yet we are forced to provide separate probabilities of “X died accidentally” given each of the 64 possible combinations of values for the other causal factors – none of which is relevant.
- *The CPT for child nodes of S* : Since most of these are also only relevant for a single hypothesis, we again have to unnecessarily specify separate probabilities conditioned on each of the different hypotheses states.

Hence, the problem we wish to solve can be characterised as follows:

*The states of the main node S correspond – by definition – to mutually exclusive alternative ‘events’ or ‘states of the world’. These separate events have independent causal parental pathways (by which we mean no common ancestors) as well as largely independent causal effects, diagnostics and evidence. Yet, because the separate events are part of a **single BN node** we are unable to disentangle the separate causes and effects. Ideally, we would like to use instead a model in which the separate ‘events’ are modelled as separate Boolean nodes as shown in Figure 2 and in which the revised model successfully preserves the prior probabilities of each of the mutually exclusive events occurring.*

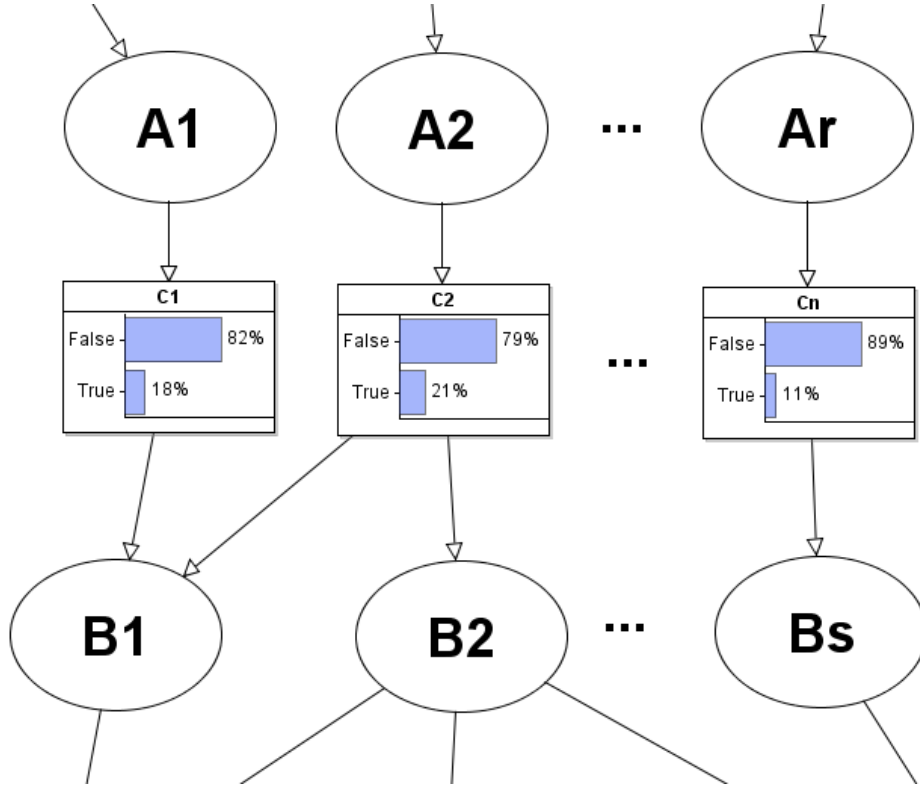


Figure 2 Ideal structure separating the mutually exclusive outcomes into distinct (Boolean) nodes (note that, although we allow common descendants of the status nodes, we do not consider common ancestors)

Specifically, and completely generally, we want to be able to define a transformation of an n -state node into n Boolean nodes, where C_i is the random (binary) variable associated with i th 'new' Boolean node. The revised model must satisfy the following two properties to ensure it is semantically equivalent to the original:

Property 1 (Basic Mutual Exclusivity): $P(C_j = \text{false} \mid C_i = \text{true}) = 1$ for each $i \neq j$

Property 2 (Equivalence of prior probabilities of the events): For each i , the prior marginal probability $P(C_i = \text{true})$ is equal to $P(S = c_i)$ in the original model.

3 PREVIOUSLY PROPOSED SOLUTIONS AND THEIR LIMITATIONS

3.1 Basic solution

There is a simple solution to the problem in the special case when there are no ancestors of the node S . In this case, the solution is to retain the node S and introduce the C_i nodes as children as shown in Figure 3.

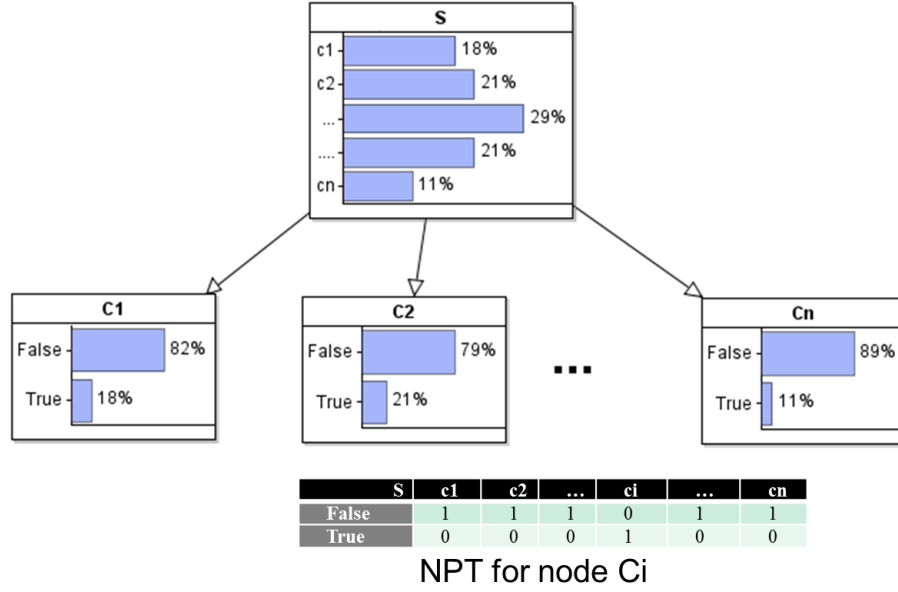


Figure 3 Solution for special case when node S has no ancestors

In this case, for $i=1$ to n , the CPT for node C_i is defined by:

$$P(C_i = \text{true} | c_j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \text{ for } j = 1, \dots, n.$$

This CPT assignment ensures both properties 1 and 2 hold:

Property 1 holds since If C_i is *true* then $P(C_j = \text{false}) = 1$ for each $i \neq j$.

Property 2 holds since

$$P(C_i = \text{true}) = \sum_{j=1}^n P(C_i = \text{true} | c_j) P(c_j) = P(c_i).$$

Clearly, because this solution involves node S as a parent of each of the consequence nodes C_i it does not help us in the case where S has ancestors.

3.2 Solution with linked C_i nodes

It turns out (as explained in Appendix B) that for the general case it *is* possible to construct a ‘solution’ that satisfies both Properties 1 and 2 by directly linking C_i nodes together. However, by introducing direct links between the C_i nodes we destroy the entire rationale for introducing separate event nodes, which was to separate the causal pathways to and from the events. Appendix B, therefore, also explains in detail why linking consequence nodes is unsatisfactory (in the solution we propose in Section 4 **there are no direct dependencies** between any of the C_i nodes). The solution in Appendix B is also infeasible where the number of states n is large, since for $i=2 \dots n$, even if the node C_1 has no causal parent nodes, it has $i-1$ enforced parent nodes (namely C_1, \dots, C_{i-1}) and hence 2^i CPT entries.

3.3 The simple constraint method

Jensen and Nielsen (2007) proposed a solution by introducing a Boolean constraint node (as shown in Figure 4 for the simplest case where $n = 2$) and setting it to be *true*. The CPT for the constraint node is a deterministic *XOR*, i.e. is defined as *true* when exactly one of the parents is *true* and *false* otherwise (so this easily generalizes to arbitrary n nodes). Providing the constraint is always set to be *true* when the model is run, Property 1 is clearly satisfied because if C_i is *true* then, since the constraint is *true*, C_j must be *false* for each $j \neq i$ because of the definition of the CPT.

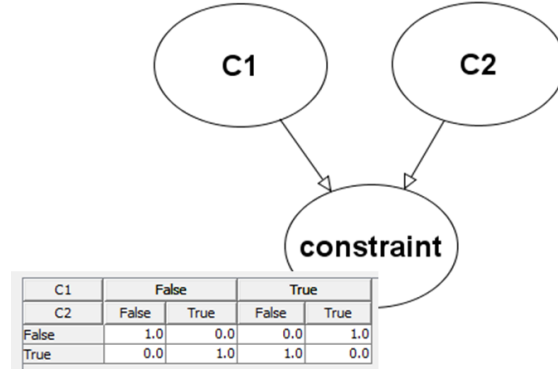


Figure 4 Enforcing mutual exclusivity by introducing simple Boolean constraint node

However, since this solution requires the constraint node to be true it does not in general preserve the prior probabilities of C_1 and C_2 , and thus it does **not** satisfy Property 2. To see this, suppose $P(C_1 = \text{true}) = x$. Then, since there are just two mutually exclusive causes, this means that we would expect $P(C_2 = \text{true}) = 1-x$. But then

$$\begin{aligned}
 P(C_1 = \text{true} | \text{Constraint} = \text{true}) &= \frac{P(C_1 = \text{true}) \times (1 - P(C_2 = \text{true}))}{(P(C_1 = \text{true}) \times (1 - P(C_2 = \text{true})) + P(C_2 = \text{true}) \times (1 - P(C_1 = \text{true})))} \\
 &= \frac{x^2}{x^2 + (1-x)^2},
 \end{aligned}$$

which is equal to x only when $x=1$ or $x=0.5$

For example, suppose $P(C_1 = \text{true}) = 0.7$. Then, $P(C_1 = \text{true} | \text{Constraint} = \text{true}) = 0.8448$. So, when the constraint is set to true (as is necessary) the priors for the cause nodes change even though no actual evidence has been entered.

It is important to note that in the examples in (Jensen and Nielsen, 2007) the priors for the mutually exclusive nodes were assumed to be uniform (i.e. in the 2-node example the prior *true* and *false* probabilities were 0.5 for each C_i node).

3.4 Extended XOR solution

Diez and Druzdzel (2006) also introduced an XOR constraint node and proposed a method that also satisfies Property 2. This solution involves directly changing the CPTs

of every cause node. Specifically, assuming the original prior probability $P(C_i=True)=x_i$ then the CPT of C_i is changed to

$$P(C_i = True) = \frac{x_i}{1 + x_i}$$

These prior probability values ensure that the posterior probability of C_i is equal to x_i when the deterministic XOR constraint is instantiated. The drawback here is that user sees a completely different set of CPTs for the original cause nodes so the model becomes unrecognisable from the original model. Furthermore, in its current form this solution also has limited value when mutually exclusive events have ancestors, as it becomes difficult to define the correct CPT values that satisfy the target posterior probability distribution.

4 PROPOSED GENERAL SOLUTION

Our solution (see Figure 5 for the structure) avoids the problems of the previous solutions and leaves all of the original BN nodes – and their relationships – unchanged. The solution is to add two nodes to the network in Figure 2: an auxiliary classification node, which is a common child of the C_i nodes, and a constraint node child of this auxiliary node to satisfy Properties 1 and 2. Setting the constraint node to true does not change the prior probabilities for the C_i nodes.

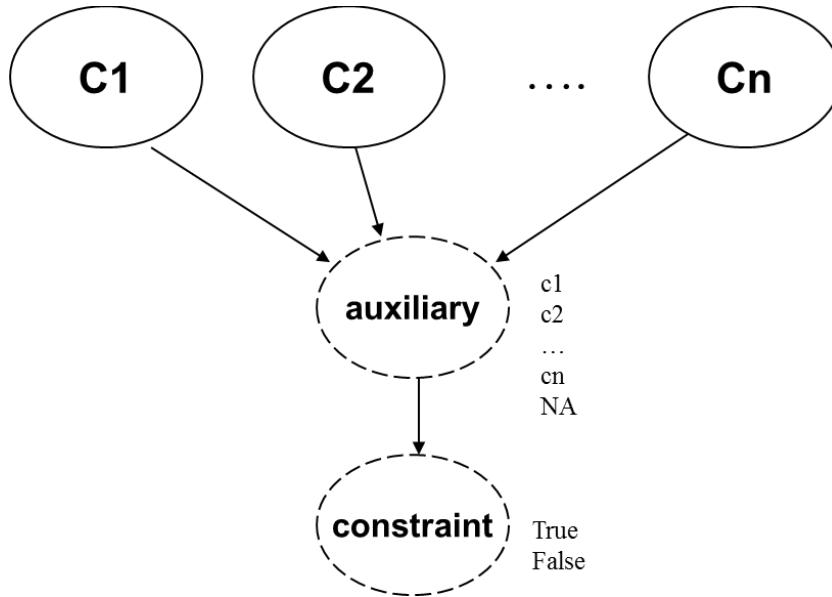


Figure 5 Structure for general solution

The auxiliary node has $n + 1$ states, namely the n original states c_i (for $i = 1$ to n) of the node S plus a special NA state standing for “Not Applicable” and representing impossible combinations. In what follows we will assume that in the original model

$P(S = c_i)=x_i$ for $i = 1$ to n and hence that the necessary prior probabilities for each node C_i are $P(C_i=true)=x_i$.

Theorem. If the CPT of the auxiliary node is defined as in Table 1, and the CPT of the constraint node is defined as in Table 2, then both properties 1 and 2 are satisfied when

the constraint is true.

Table 1 CPT for auxiliary cause node (specifically:

$P(c_i = 1)$ when $C_i = \text{true}$ and $C_j = \text{false}$ for each $i \neq j$; otherwise $P(c_i = 0)$;

$P(NA) = 0$ if exactly one C_i is true and 1 otherwise)

C_1	<i>False</i>				<i>True</i>			
C_2	<i>False</i>		<i>True</i>		<i>False</i>		<i>True</i>	
...								
C_n	<i>False</i>	<i>True</i>	<i>False</i>	<i>True</i>	<i>False</i>	<i>True</i>	<i>False</i>	<i>True</i>
c_1	0	0	0	0	1	0	0	0
c_2	0	0	1	0	0	0	0	0
...								
c_n	0	1	0	0	0	0	0	0
NA	0	0	0	1	0	1	1	1

Table 2 CPT for constraint node

<i>Auxiliary</i>	c_1	c_2	...	c_n	NA
<i>False</i>	x_1	x_2	...	x_n	1
<i>True</i>	$1 - x_1$	$1 - x_2$...	$1 - x_n$	0

Proof

First we prove Property 1 holds, i.e. $P(C_j = \text{false} \mid C_i = \text{true}) = 1$ for each $i \neq j$.

We argue by contradiction. Suppose Property 1 does **not** hold. Then C_j is *true* for some $i \neq j$ when C_i is *true*. But, if C_i and C_j are both true then, from the definition of the CPT for the auxiliary node, we must have $P(NA) = 1$. But then, from the definition of the CPT for the constraint node $P(\text{constraint} = \text{true}) = 0$, which contradicts the fact that the constraint is ‘true’.

To prove property 2 we have to show that the marginal probabilities for the C_i nodes do not change when the constraint is set to true, i.e. we have to show for each $i = 1$ to n

$$P(C_i = \text{true} \mid \text{constraint} = \text{true}) = P(C_i = \text{true}) \quad (1)$$

By Bayes

$$P(C_i = \text{true} \mid \text{constraint} = \text{true}) = \frac{P(\text{constraint} = \text{true} \mid C_i = \text{true}) \times P(C_i = \text{true})}{P(\text{constraint} = \text{true})} \quad (2)$$

Hence, if we can show

$$P(\text{constraint} = \text{true} \mid C_i = \text{true}) = P(\text{constraint} = \text{true}), \quad (3)$$

it follows from (2) that (1) is true.

When $C_i = \text{true}$ it follows from the definition of the CPT for the auxiliary node that

$$P(c_j|C_i = true) = \begin{cases} \prod_{j \neq i}^n P(C_j = false) = \prod_{j \neq i}^n (1 - x_j) & \text{if } j = i \\ 0 & \text{if } j \neq i \end{cases} \quad (4)$$

Using (4) it follows from the definition of the CPT for the constraint node and marginalisation that:

$$\begin{aligned} P(constraint = true|C_i = true) &= \sum_{j=1}^n (1 - x_j) \times P(c_j|C_i = true) \\ &= (1 - x_i) \times \prod_{j \neq i}^n (1 - x_j) \quad (\text{by (4)}) \\ &= \prod_{i=1}^n (1 - x_i) \quad (5) \end{aligned}$$

Now we know from the definition of the CPT for the auxiliary node that:

$$\begin{aligned} P(c_1) &= P(C_1 = true, C_2 = false, \dots, C_n = false) = x_1(1 - x_2) \dots (1 - x_n) \\ P(c_2) &= P(C_1 = false, C_2 = true, \dots, C_n = false) = (1 - x_1)x_2 \dots (1 - x_n) \\ &\dots \\ P(c_n) &= P(C_1 = false, C_2 = false, \dots, C_n = true) = (1 - x_1)(1 - x_2) \dots x_n \end{aligned}$$

So, in general for each i :

$$P(c_i) = x_i \prod_{j \neq i}^n (1 - x_j)$$

Using this together with the definition of the CPT for the constraint node and marginalisation:

$$\begin{aligned} P(constraint = true) &= \sum_{i=1}^n (1 - x_i) P(c_i) \\ &= \sum_{i=1}^n \left((1 - x_i) x_i \prod_{j \neq i}^n (1 - x_j) \right) \\ &= \sum_{i=1}^n \left(x_i \prod_{i=1}^n (1 - x_i) \right) \\ &= \prod_{i=1}^n (1 - x_i) \left(\sum_{i=1}^n x_i \right) \\ &= \prod_{i=1}^n (1 - x_i), \quad (6) \end{aligned}$$

since $\sum_{i=1}^n x_i = 1$

This completes the proof. An example using the solution is shown in Figure 6.

There are four important points to note about the solution:

1. *The values in Table 2 are not unique.* In fact, it follows from the above proof that any constant multiple of the values will also work (provided the results are all between 0 and 1 as they are probabilities assigned to a CPT), i.e. for any constant, for which

$$0 < k(1 - x_i) < 1 \text{ for each } i.$$

Multiplying the probabilities by k is correct because only the relative likelihoods transmitted from ‘constraint’ to ‘auxiliary’ are relevant. For example, Table 3 also works.

Table 3 alternative CPT for constraint node with constant multiple

<i>Auxiliary</i>	c_1	c_2	...	c_n	<i>NA</i>
<i>False</i>	$1-k(1-x_1)$	$1-k(1-x_2)$...	$1-k(1-x_n)$	1
<i>True</i>	$k(1-x_1)$	$k(1-x_2)$...	$k(1-x_n)$	0

2. *It extends to non-exhaustive events.* If our starting point for the mutual exclusivity problem is a node whose states we wish to represent as separate nodes then, by definition, the set of states are not only mutually exclusive but also *exhaustive*. However, in many situations our starting point for the problem is a BN in which we already have constructed separate nodes that we wish to force to be ‘mutually exclusive’. In such situations the set of states may not be exhaustive. The proposed solution works in such situations by simply adding a ‘leak’ state to the auxiliary node. This state represents the logical alternative

$$\text{not}(C_1 \text{ or } C_2 \text{ or } \dots \text{ or } C_n),$$

where C_1, \dots, C_n are the nodes representing the known mutually exclusive events. By definition adding this state ensures the set $\{C_1, \dots, C_n, \text{leak}\}$ is mutually exclusive and exhaustive, and the prior probability of the state *leak* is simply $1 - \sum_{i=1}^n x_i$ where x_i is the prior probability of state C_i . Hence, the necessary CPT for the constraint node is that shown in Table 4.

Table 4 CPT for constraint node where the states C_1, \dots, C_n are not exhaustive

<i>Auxiliary</i>	c_1	c_2	...	c_n	<i>leak</i>	<i>NA</i>
<i>False</i>	x_1	x_2	...	x_n	$\sum_{i=1}^n x_i$	1
<i>True</i>	$1-x_1$	$1-x_2$...	$1-x_n$	$1 - \sum_{i=1}^n x_i$	0

3. *There is an equivalent solution that does not require a separate constraint node.* The constraint on the auxiliary node can also be imposed by using *virtual evidence* (Bilmes, 2004, Butz and Fang, 2005, Chan and Darwiche, 2005) directly on the auxiliary. This is explained in Appendix C, but this solution is more difficult to implement practically.
4. *The solution does not work if there are common ancestors of the C_i nodes.*

Suppose, for example, that two nodes C_i and C_j have a common parent node A . Then these nodes are d-connected once the evidence “true” is entered in the constraint node. So back propagation from C_i to A can result in a change to $P(C_j)$. No such change is possible if there are no common ancestors. There are no problems for common descendants since evidence in the constraint node does not change the conditional independence assertions of descendant nodes.

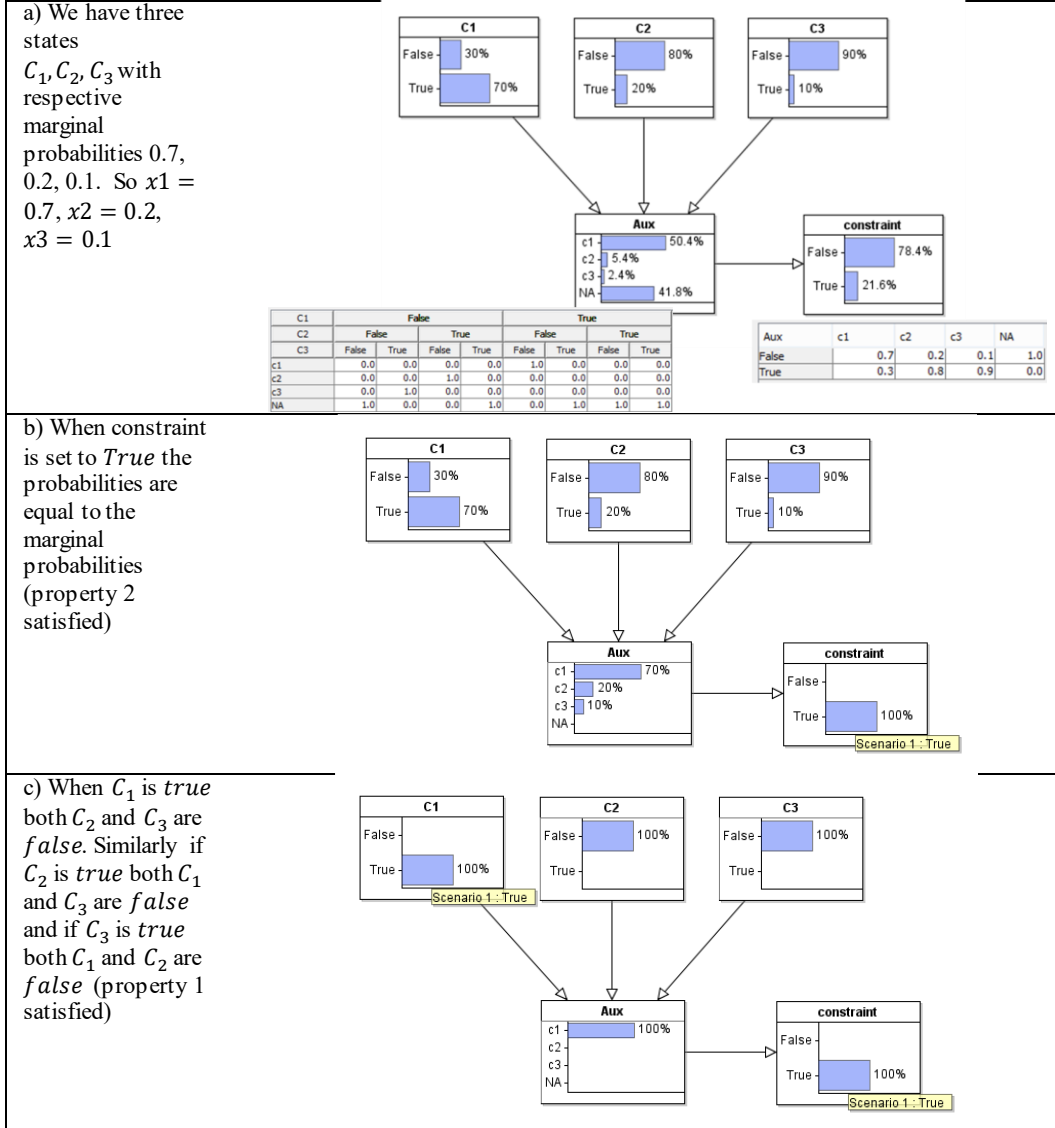


Figure 6 Example model showing the solution in action ($n = 3$)

5 USING THE PROPOSED SOLUTION IN PRACTICE

Our original motivation for solving the problem was driven by its frequent occurrence in handling legal arguments where the prosecution and defence hypotheses normally have clearly different causal explanations and evidence (Fenton et al., 2013). Indeed, in that paper we identified it as an ‘idiom’ that was a special case of the classic ‘explaining away’ reasoning. Figure 7 shows an example fragment of a model from (Fenton et al., 2013) in which two nodes need to be separate (because of different causal pathways) but also have to be mutually exclusive. The solution proposed in Section 4 achieves the logically ‘correct’ probabilistic reasoning in this model when evidence is

entered. Ordinary ‘explaining away’ (which can only be achieved by the definition of the CPT for node ‘Blood on shirt matches victim blood’) simply does not work.

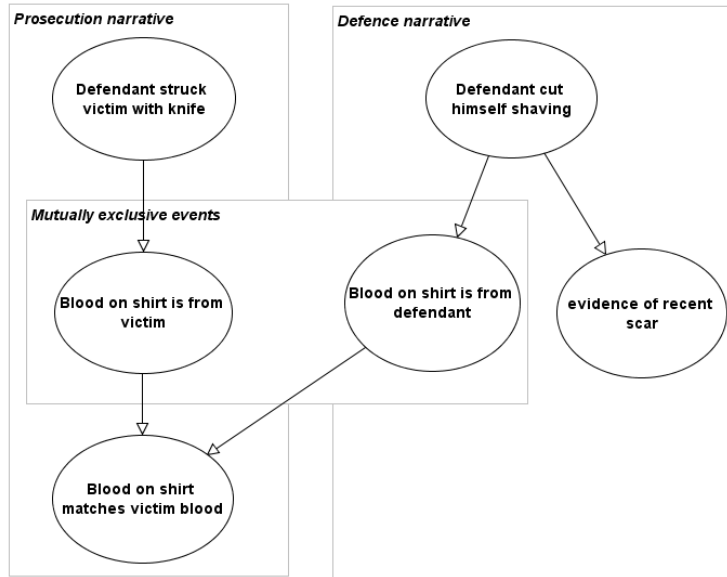


Figure 7 Blood is found on the defendant’s shirt

Our proposed solution has also already been adopted by Vlek et al. (2014) to analyse evidence in legal trials using narratives. Vlek et al. (2014) used mutually exclusive causes to combine multiple BN fragments about different narratives in a real legal case study. Each BN fragment represented an alternative scenario explaining the available evidence. Since only one scenario can be true, the scenarios are modelled as mutually exclusive causes of the incident.

The proposed method comes with some limitations:

- The benefits of the method are limited when the number of mutually exclusive causes is low. In real-world problems, however, it is very common to deal with multiple causes and/or classifications, and in these cases the method becomes useful.
- There is a need to update the CPT of the constraint node when the CPTs of the ancestors of the C_i ’s change. Care is needed in situations where the mutually exclusive events represent alternative ‘causes’ of some incident.
- It may be impossible to identify the set of all potential causes and hence there is a danger in assuming that the set of identified potential causes is exhaustive. This danger is especially pertinent for some legal arguments, where the causes represent different hypotheses (for example, the defendant fired a weapon without provocation or fired in self-defence). Strong evidence against one hypothesis here would result in favouring the other. This would be an inappropriate conclusion in the case where an additional potential hypothesis, say ‘fire by accident’, had been wrongly omitted from the model. The ‘fix’ to this problem is either to accept that the causes are not exhaustive (and to use probability to deal with unknown causes) or to add a catch-all ‘other/unknown’ cause to the list of known causes. The *none* state in our solution represents the

case where all causes in the model are *false* and therefore it can be used to model ‘other/unknown’ causes that are not included in the model. However, modelling unknown causes in a BN model creates different problems, notably that of completing the necessary prior conditional probabilities for the effect given a cause that we do not know.

6 CONCLUSIONS

BNs have proven to be a very powerful method for reasoning about uncertainty. However, in situations where we wish to model mutually exclusive events as part of a complex argument there is no ‘natural’ BN model that works. We can:

- *Model the events as the states of a single node.* But this comes at the heavy cost of introducing complex and often meaningless sets of conditional probabilities. In practice we have found this ‘solution’ is unsatisfactory and often infeasible.
- *Introduce direct links between the nodes.* However, the ‘obvious’ solution fails to retain the prior probabilities of the event states. Although we have shown it is possible to get round this problem using a special assignment of CPT values (Appendix B), any linked nodes solution defeats the objective of keeping causal pathways into individual nodes separate. Hence, this ‘solution’ never makes sense unless the event nodes have no parents.

We showed that the proposed ‘solution’ (Section 3.3) of introducing a simple XOR constraint node and setting it to ‘*true*’ fails to retain the prior probabilities of the event states except in special cases. The proposed extended XOR solution (Section 3.4) does preserve the priors but at the unacceptably heavy cost of having to redefine every event node CPT. The verified solution we have proposed makes no changes at all to the original BN nodes and their links. What we have done is produce an auxiliary node and a constraint node that is set to ‘*true*’. Our solution provides a simple method for assigning the necessary CPT values in all cases to these new nodes.

Our solution is by no means ideal since it involves additional artificial nodes, as well as a *NA* state that represents impossible state combinations, and a *None* state where the mutually exclusive events are not exhaustive. However, there is no solution to the mutual exclusivity problem without changing the original BN structure, and the solution we have provided solves the problem under the properties described in this paper, which represent common modelling scenarios.

Although we have described the problem and its solution in generic form this is certainly not a purely theoretical exercise. Examples mentioned in the paper (legal arguments and military threat assessment) are examples involving real clients where we needed to model mutually exclusive causes and discovered that the standard BN solutions did not work.

ACKNOWLEDGMENT

This is supported by ERC project ERC-2013-AdG339182-BAYES_KNOWLEDGE. The work has benefited from the input of a number of colleagues and experts. We are also indebted to Agena for use of their software.

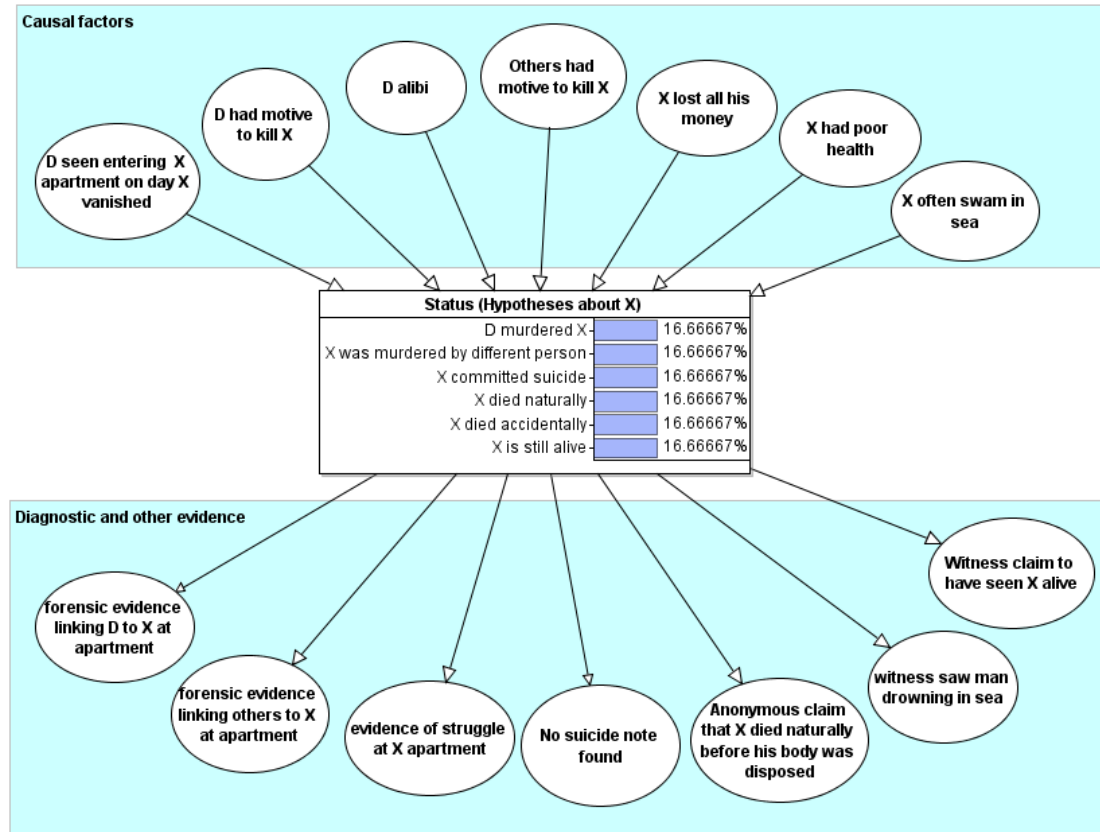
APPENDIX A: EXAMPLE PROBLEM

Figure 8 Simplified example of BN model used for prognosis and diagnosis

The example BN in Figure 8 is a simplified version of a typical BN used to model legal arguments. In this example a defendant D is charged with the murder of X. Although the body of X has never been discovered the prosecution case is based around the following evidence:

- **Opportunity and motive (causal factors):** A witness claims to have seen D entering X's apartment the day before he was reported missing (opportunity) and other witnesses claim that D had said he wanted to kill X following a dispute (motive).
- **Diagnostic and other evidence:** Various forensic evidence found on D and at X's apartment after X's disappearance linking D to a violent confrontation with X; evidence of a struggle taking place at X's apartment

The defence has one piece of evidence (an alibi from D that challenges the 'opportunity' evidence of the prosecution) but also suggests a number of possible alternatives hypotheses to the prosecution's each of which has its own different narrative and evidence (both causal and diagnostic). The ideal structure for this problem, involving separate 'mutually exclusive' nodes for each hypothesis about X is shown in Figure 9.

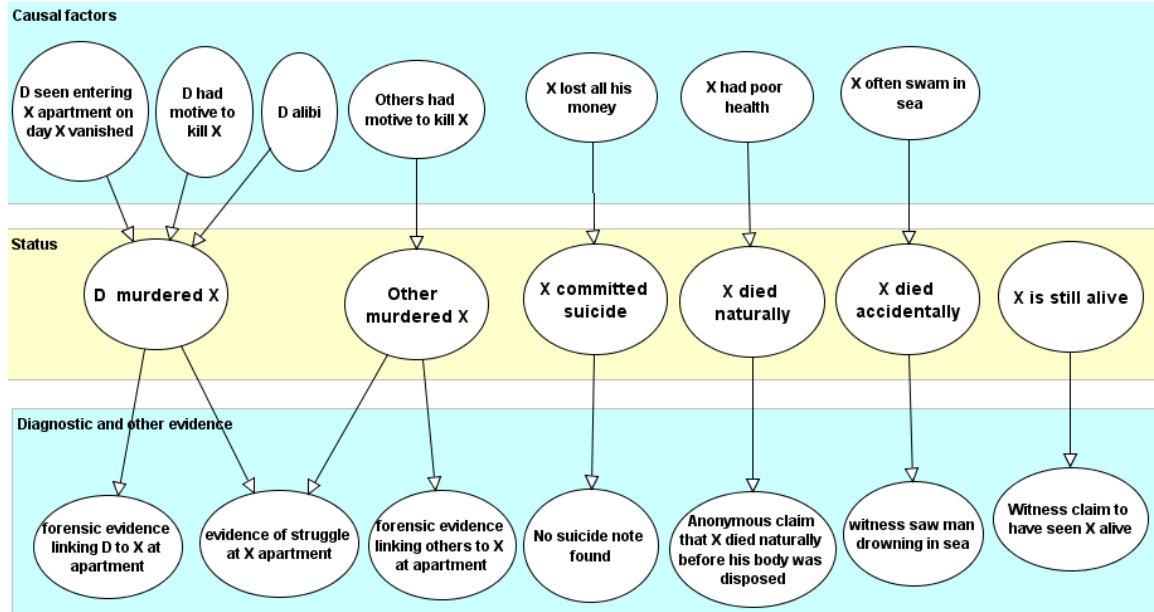


Figure 9 Ideal structure separating the mutually exclusive outcomes into distinct (Boolean) nodes (note that, although we allow common descendants of the status nodes, we do not consider common ancestors)

APPENDIX B: SOLUTION INVOLVING DIRECT DEPENDENCIES BETWEEN THE C_i NODES

Mutual exclusivity between separate nodes can also be enforced by adding edges between every pair of those nodes. The main task in this case is to assign the CPTs of those in such a way that both properties of mutual exclusivity are satisfied, and to avoid introducing cycles to the BN structure. Let C_1, \dots, C_n be separate nodes that we want to enforce mutual exclusivity. In order to have edges between every pair of those nodes, we need to add $\binom{n}{2}$ edges.

Suppose any edge added between C_i to C_j is directed from C_i to C_j where $i < j$. In other words, C_1 has no parents, C_2 has 1 parent (i.e. C_1), C_n has $n - 1$ parents (i.e. C_1, \dots, C_{n-1}) etc. (Figure 10 shows an example where $n = 3$). Our task is to assign the CPTs of C_1, \dots, C_n in such a way that both properties of mutual exclusivity are satisfied. A possible ‘solution’ is to define the CPT of the node C_i (where $i > 1$) to be *false* for all columns except one: the column in which all the parents are *false*. For this column, the CPT is defined as:

- for $i < n$: *true* and *false* are both assigned probabilities 0.5
- for n : it must be *true* (i.e. *true* is assigned probability 1)

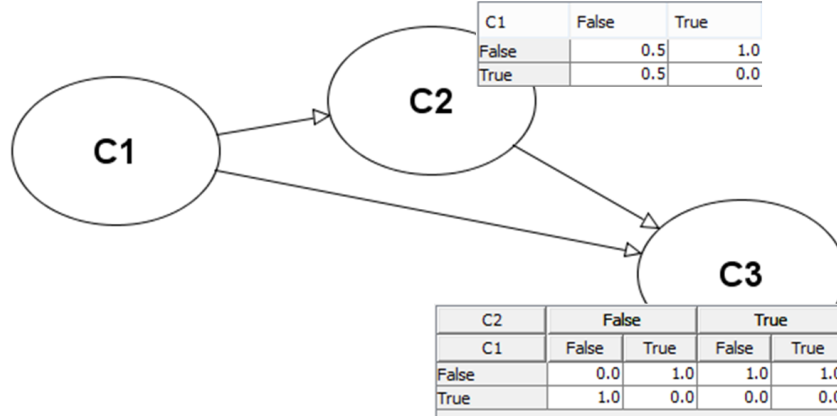


Figure 10 Ensuring mutual exclusivity through direct dependencies between the C_i nodes. Note that the last column of the CPT for C3 represents an impossible state combination and so can be defined with any values.

However, it turns out that the solution *fails to satisfy property 2* (equivalence of prior marginals), except in some special cases (such as when $n = 2$). This is because whatever value x_1 , i.e. the prior of C_1 , is set to the other C_i 's will have the following marginal values for *true*:

In the case where $n = 3$

$$P(C_2 = \text{True}) = (1 - x_1)/2$$

$$P(C_3 = \text{True}) = (1 - x_1)/2$$

In the case where $n = 4$:

$$P(C_2 = \text{True}) = (1 - x_1)/2$$

$$P(C_3 = \text{True}) = (1 - x_1)/4$$

$$P(C_4 = \text{True}) = (1 - x_1)/4$$

In the case where $n = 5$:

$$P(C_2 = \text{True}) = (1 - x_1)/2$$

$$P(C_3 = \text{True}) = (1 - x_1)/4$$

$$P(C_4 = \text{True}) = (1 - x_1)/8$$

$$P(C_5 = \text{True}) = (1 - x_1)/8$$

etc.

So, in the special case when $n = 3$ and each of the priors happens to be $1/3$ the solution will work as shown in Figure 11.

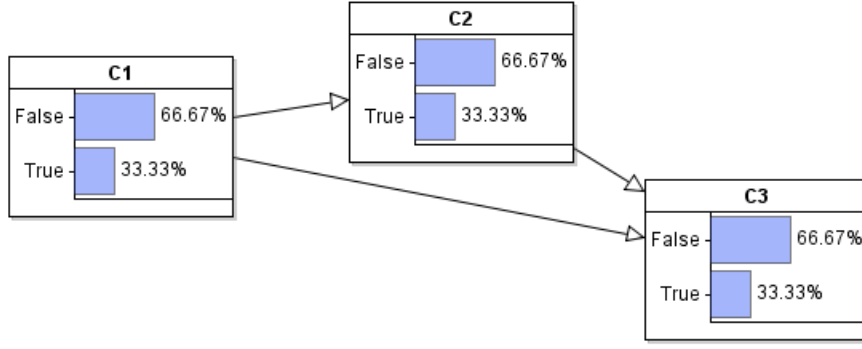


Figure 11 Marginal values for $C_i = \text{true}$ are all equal to $1/3$

However, suppose that the marginals for states c_1, c_2, c_3 are 0.7, 0.2 and 0.1 respectively. Then, because $x_2 \neq x_3$, the marginals are not preserved (as shown in Figure 12).

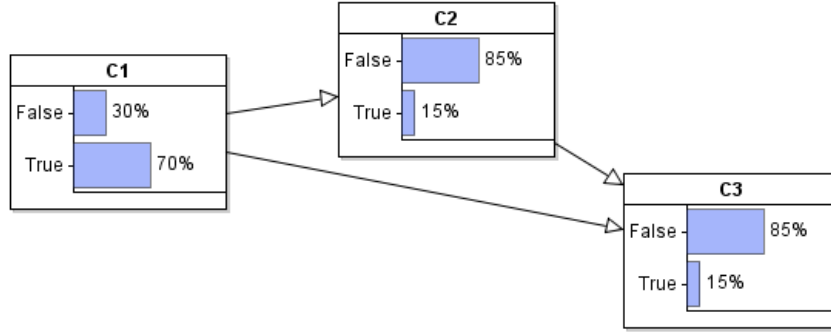


Figure 12 Marginals of 0.7, 0.2, 0.1 are NOT preserved.

In order to preserve the marginals we have to be much more careful in the definition of the CPTs of nodes C_i where $1 < i < n$. Specifically, we cannot assign uniform values to *true* and *false* for the column where all parents are *false*.

Instead, we have to assign values that preserve the priors. Now we know that the marginal probability $P(C_i = \text{true})$ is simply the sum of all probabilities of the form:

$$P(C_i = \text{true} | C_1 = \text{true}, C_2 = \text{false}, \dots, C_{i-1} = \text{true}) P(C_1 = \text{true}) \dots P(C_{i-1} = \text{true})$$

Where we consider all state combinations of the parents C_1, \dots, C_{i-1}

However, we also know that the conditional probability that C_i is *true* given the parents' states is 0 unless all the parents are *false*. Hence, we can conclude that the marginal probability $P(C_i = \text{true})$ is equal to:

$$P(C_i = \text{true} | C_1 = \text{false}, C_2 = \text{false}, \dots, C_{i-1} = \text{false}) P(C_1 = \text{false}) \dots P(C_{i-1} = \text{false})$$

However, we need $P(C_i = \text{true})$ to be equal to the marginal probability for the state c_i , i.e. x_i . Then it follows that:

$$x_i = P(C_i = \text{true} | C_1 = \text{false}, C_2 = \text{false}, \dots, C_{i-1} = \text{false})(1 - x_1)(1 - x_2) \dots (1 - x_{i-1})$$

And hence

$$P(C_i = \text{true} | C_1 = \text{false}, C_2 = \text{false}, \dots, C_{i-1} = \text{false}) = \frac{x_i}{(1 - x_1)(1 - x_2) \dots (1 - x_{i-1})}$$

So, the required CPT entry for C_i being *true* when all the parents are *false* is:

$$\frac{x_i}{(1 - x_1)(1 - x_2) \dots (1 - x_{i-1})}$$

As an example consider the case where the marginals for c_1, c_2, c_3 are respectively 0.7, 0.2, 0.1 then it is the CPT for node C_2 that has to be redefined with $P(C_2 = \text{true} | C_1 = \text{false})$ equal to $0.2/(1 - 0.7) = 2/3$. With this assignment we get the preserved marginal as shown in Figure 13.

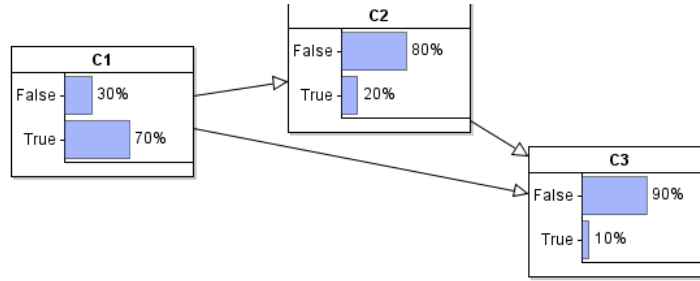


Figure 13: Marginals 0.7, 0.2, 0.1 are preserved

Although we have shown that it is possible to configure the CPTs such that properties 1 and 2 are both satisfied, the solution is unsatisfactory because it compromises the fact that the main objective of creating separate nodes for each outcome was to separate the largely independent causal pathways to and from the outcomes. By introducing direct links between the C_i nodes we destroy the separation, and actually create a model that is far more complex than the original. Not only do we now need carefully constructed CPTs for each of the C_i nodes conditioned on other C_j nodes, but these CPTs have to be completely redefined as soon as there are causal parents for node C_i . We are again forced to consider all the irrelevant combinations of states of all the other C_j nodes in defining the CPT for C_j given the causal factors. Even without causal factor parents, the CPTs involve a whole range of meaningless columns and impossible state combinations.

Another unsatisfactory aspect of this ‘solution’ is the fact that we have to arbitrarily decide which one of C_i and C_j is to be the ‘parent’ of the other even in cases where a temporal or causal association between these mutually exclusive events simply does not exist.

APPENDIX C: VIRTUAL EVIDENCE SOLUTION

The constraint node linked to the auxiliary node (described in Section 4) can also be interpreted as uncertain evidence and handled accordingly. BNs have two types of uncertain evidence – virtual and soft evidence – that are often confused with each other (Pearl, 1988, Bilmes, 2004, Chan and Darwiche, 2005). Virtual evidence uses a likelihood ratio to represent the uncertainty of evidence. The formal definition of virtual evidence is as follows:

Let η be some uncertain evidence imposed on a set of mutually exclusive and exhaustive events c_1, \dots, c_n and assume that such evidence is specified by w_1, \dots, w_n such that:

$$P(\eta|c_1) : \dots : P(\eta|c_n) = w_1 : \dots : w_n,$$

the revised distribution proposed by the virtual evidence is $P(.|\eta)$. The virtual event η is independent of all other events given c_i for $i=1, \dots, n$.

In BNs, virtual evidence can be manually modelled by:

1. adding a dummy node that corresponds to η ,
2. adding a directed edge between the node containing c_i and the dummy node,
3. instantiating the dummy node.

The dummy node is equivalent to the constraint node in our method. The conditional probabilities of the instantiated state of the dummy node are defined according to the likelihood ratio of uncertain evidence.

Virtual evidence is implemented in many commercial BN software packages that automatically handle the dummy variable and its CPT. When we use such software, our task is only to set the virtual evidence weights representing likelihood ratios:

$$(w_1, w_2, \dots, w_n, w_{NA})$$

for the respective states c_1, \dots, c_n, NA in such a way that the resulting marginal for the auxiliary node is equal to:

$$(x_1, x_2, \dots, x_n, 0).$$

Note that we also have to define a weight for the NA state in our auxiliary node. The required virtual evidence weights are the same as the CPT parameters of the constraint node described in Section 4:

$$w_i = k(1 - x_i) \text{ for } i = 1, \dots, n,$$

$$w_{NA} = 0,$$

where k is any constant for which $1 > k(1 - x_i) > 0$.

Figure 14 shows how the constraint is imposed with virtual evidence in AgenaRisk (2016) using the same example shown in Section 4.

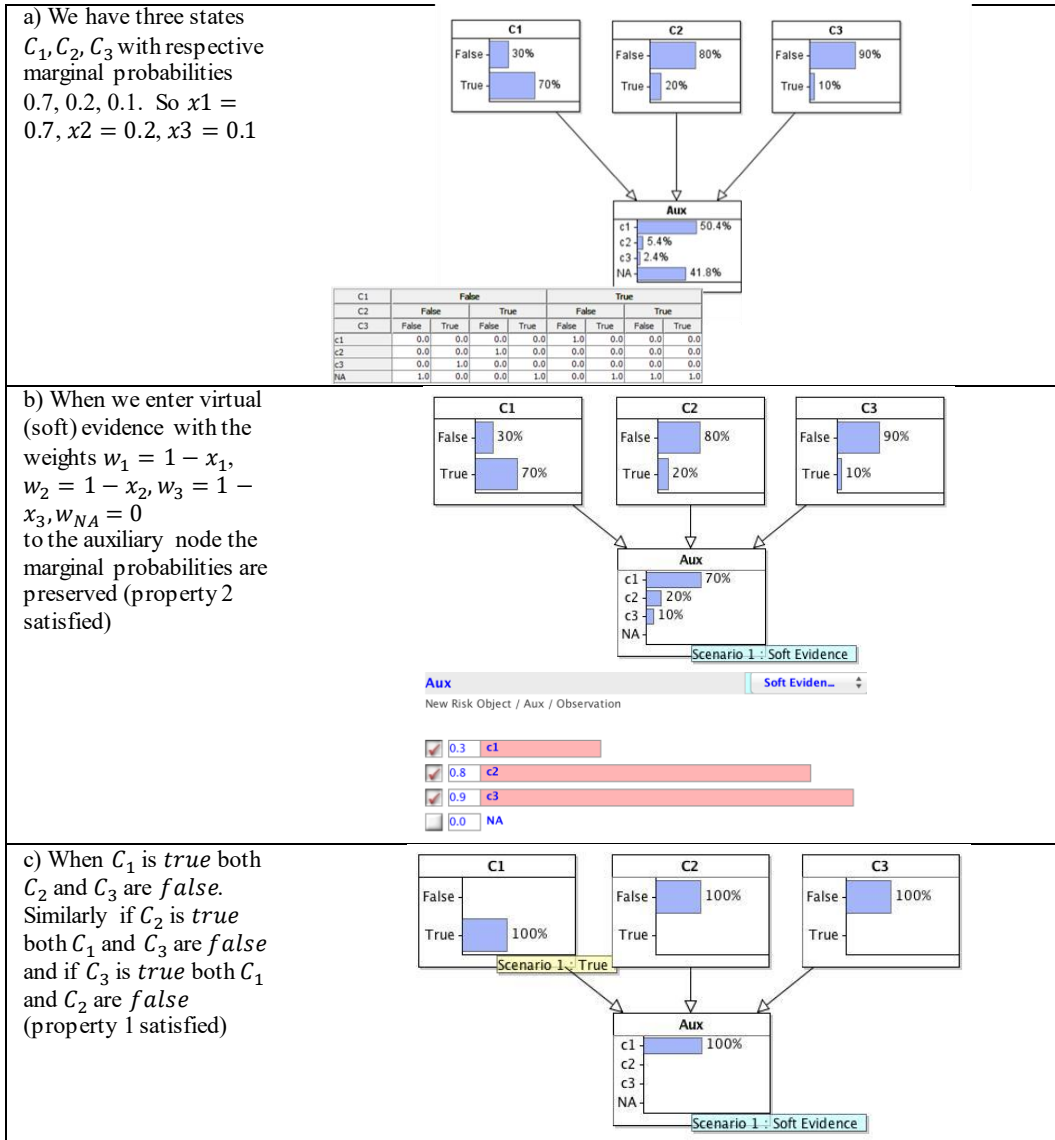


Figure 14 Alternative Solution using Virtual (Soft) Evidence

We use virtual evidence to set the probabilities of the auxiliary node to a target value. This is analogous to the second type of uncertain evidence called soft evidence. While virtual evidence uses uncertain evidence values as likelihood ratios, soft evidence uses them as the target posterior distribution. In other words, when soft evidence is applied, the result of propagating the BN is such that the marginal distribution for the node N is exactly the array (x_1, x_2, \dots, x_n) , where x_i 's sum to one. The formal definition of soft evidence is as follows:

Let η be some uncertain evidence imposed on a set of mutually exclusive and exhaustive events c_1, \dots, c_n , the revised distribution proposed by soft evidence satisfy the following constraint

$$P(c_i|\eta) = x_i.$$

The solution described in this appendix uses the virtual evidence updating to produce the soft evidence outcome. Another way to impose soft evidence by using virtual evidence is to use the following likelihood ratios as weights of the virtual evidence.

$$\frac{x_1}{p_1} : \frac{x_2}{p_2} : \dots : \frac{x_n}{p_n},$$

where p_i are the probabilities of the states c_i before the uncertain evidence is applied for $i=1, \dots, n$. In our example, p_1, p_2, p_3 and p_{NA} are the probabilities in the auxiliary node before the constraint is imposed, and their values are 0.504, 0.054, 0.024 and 0.418 respectively (see Figure 14). The target probabilities of the states of the auxiliary nodes are 0.7, 0.2, 0.1 and 0 respectively. The likelihood ratios that satisfy these target probabilities are:

$$\frac{0.7}{0.504} : \frac{0.2}{0.054} : \frac{0.1}{0.024} : \frac{0}{0.418}.$$

These likelihood ratios are the equivalent to the virtual evidence weights computed by $w_i = k(1 - x_i)$.

Soft evidence is not readily implemented in the popular BN tools; the likelihood ratios that satisfy the desired soft evidence must be manually calculated and entered by users (Murphy, 2014). There are propagation algorithms for soft evidence (for example ‘big clique’ (Kim et al., 2004), IPFP (Pan et al., 2006) but none are provided with any of the widely available BN tools. There is also a philosophical concern about whether soft evidence has any rational meaning in the real world. (Pearl, 1988) considered that the only sensible type of ‘uncertain evidence’ that could rationally be specified was ‘virtual evidence’. The distinction and comparison between the different types of uncertain evidence is explained in detail by (Chan and Darwiche, 2005, Bilmes, 2004, Pearl, 1988).

REFERENCES

- AGENARISK. 2016. *AgendaRisk: Bayesian network and simulation software for risk analysis and decision support* [Online]. Available: <http://www.agenarisk.com> [Accessed 19/07/2016].
- BILMES, J. 2004. On virtual evidence and soft evidence in bayesian networks. Technical Report, UWEETR-2004-0016, University of Washington.
- BUTZ, C. J. & FANG, F. Incorporating evidence in Bayesian networks with the select operator. Conference of the Canadian Society for Computational Studies of Intelligence, 2005. Springer, 297-301.
- CHAN, H. & DARWICHE, A. 2005. On the revision of probabilistic beliefs using uncertain evidence. *Artificial Intelligence*, 163, 67-90.
- DIEZ, F. J. & DRUZDZEL, M. J. 2006. Canonical probabilistic models for knowledge engineering. Technical Report CISIAD-06-01, UNED, Madrid, Spain.
- FENTON, N. & NEIL, M. 2012. *Risk assessment and decision analysis with Bayesian networks*, CRC Press.
- FENTON, N., NEIL, M. & LAGNADO, D. A. 2013. A general structure for legal arguments about evidence using Bayesian networks. *Cognitive science*, 37, 61-102.
- FLORES, M. J., GÁMEZ, J. A. & MORAL, S. 2005. Abductive inference in Bayesian networks: finding a partition of the explanation space. *Symbolic and Quantitative Approaches to Reasoning with*

Uncertainty. Springer.

JENSEN, F. V. & NIELSEN, T. D. 2007. *Bayesian networks and decision graphs*, Springer.

KIM, Y.-G., VALTORTA, M. & VOMLEL, J. 2004. A prototypical system for soft evidential update. *Applied Intelligence*, 21, 81-97.

LAM, F. & YEAP, W. K. 1992. Bayesian updating: on the interpretation of exhaustive and mutually exclusive assumptions. *Artificial intelligence*, 53, 245-254.

MURPHY, K. P. 2014. *Software packages for graphical models* [Online]. Available: <http://www.cs.ubc.ca/~murphyk/Software/bnsoft.html> [Accessed 19/07/2016].

PAN, R., PENG, Y. & DING, Z. Belief update in Bayesian networks using uncertain evidence. Tools with Artificial Intelligence, 2006. ICTAI'06. 18th IEEE International Conference on, 2006. IEEE, 441-444.

PEARL, J. 1988. *Probabilistic reasoning in intelligent systems: networks of plausible inference*, Morgan Kaufmann.

PEARL, J. 2000. *Causality: models, reasoning and inference*, Cambridge Univ Press.

PERRY, B. B. & VAN ALLEN, T. Causal reasoning engine: An explanation-based approach to syndromic surveillance. System Sciences, 2005. HICSS'05. Proceedings of the 38th Annual Hawaii International Conference on, 2005. IEEE, 143b-143b.

SAMARAS, N. & STERGIOU, K. 2005. Binary encodings of non-binary constraint satisfaction problems: algorithms and experimental results. *Journal of Artificial Intelligence Research*, 24, 641-684.

VLEK, C. S., PRAKKEN, H., RENOOIJ, S. & VERHEIJ, B. 2014. Building Bayesian networks for legal evidence with narratives: a case study evaluation. *Artificial Intelligence and Law*, 22, 375-421.

VITAE

Norman Fenton is Professor of Computing at Queen Mary University of London and also CEO of Agena, a company that specialises in risk assessment software for critical systems. Norman has been involved in both the research and practice of software engineering for 25 years, but his recent research has focused on causal models (Bayesian Nets) for risk assessment. In addition to extensive work on causal models for software defect prediction, this work covers a wide range of application domains such as legal reasoning (he has been an expert witness in major criminal and civil cases), medical trials, vehicle reliability, embedded software, transport systems, and financial services.

Martin Neil Martin is Professor in Computer Science and Statistics at Queen Mary University of London, where he teaches decision and risk analysis and software engineering. Martin is also a joint founder and Chief Technology Officer of Agena Ltd. Martin has over twenty years experience in academic research, teaching, consulting, systems development and project management. His interests cover Bayesian modelling and/or risk quantification in diverse areas: operational risk in finance, systems and design reliability (including software), software project risk, decision support, simulation (using dynamic discretisation as an alternative to Monte Carlo) cost benefit analysis, AI and personalization, and statistical learning.

David Lagnado is Senior Lecturer in Cognitive and Decision Sciences at the Division of Psychology & Language Sciences, University College London. His research focuses on the psychological processes that underlie human learning, reasoning and decision-making. His recent work has explored the role of causal models in evidential reasoning and juror decision making.

William Marsh is a Lecturer in the School of Electronic Engineering and Computer Science at Queen Mary University of London. His research focuses on applied risk modelling and decision-support in both medicine and in systems engineering.

Barbaros Yet is an Assistant Professor in the Department of Industrial Engineering at Hacettepe University, Ankara, Turkey. He has a PhD in Computer Science from Queen Mary University of London. His research focuses on decision support and risk analysis using Bayesian and graphical probabilistic models with applications in project management and medicine.

Anthony Constantinou is Post-Doctoral researcher at Queen Mary, University of London. His research interests lie in Bayesian Artificial Intelligence theory and modelling for optimal real-world decision making. Most of his research is driven by acknowledging the limitations of classical statistics and automated learning algorithms, and inventing methods that incorporate 'causal intelligence' to the process of data engineering, and real-world 'causal facts' to the process of causal discovery, driven by what information is really required for causal representation.